3. MANUAL DEL USUARIO

3.1. INTRODUCCIÓN

Modelo de Red 2 v.1.4 es un programa que permite realizar, a partir de un archivo de texto en formato fasta, la clasificación o predicción de promotores en secuencias genéticas. La aplicación permite cargar un archivo fasta y aplicar sobre los datos leídos cinco posibles algoritmos de clasificación o predicción de promotores, donde cada algoritmo es una red neuronal artificial entrenada para determinado tipo de organismo, a saber, la rata, el ratón, el gallo, el toro y el humano. Dependiendo de la longitud de las secuencias de entrada la aplicación clasificará promotores si son de tamaño de 250 nucleótidos o predecirá si son de mayor tamaño. Además, si el archivo de entrada se organiza en determinada forma, la aplicación calculará y mostrará diferentes indicadores de exactitud. Modelo de Red 2 permite guardar los resultados obtenidos en un archivo de texto plano para su posterior análisis además de facilitar su portabilidad.

3.2. REQUERIMIENTOS BÁSICOS

Lenguaje de programación Java versión 1.6. Computador con procesador de 1GHz y 500MB de RAM.

3.3. APLICACIÓN

El usuario desde el menú "RNAs" o desde la lista desplegable seleccionará una red neuronal artificial (RNA) entrenada para determinado organismo, por defecto, modelo de Red 2 trabajará con la RNA entrenada para la especie rata. En la Figura 1 se selecciona como ejemplo la RNA entrenada para la especie gallo.

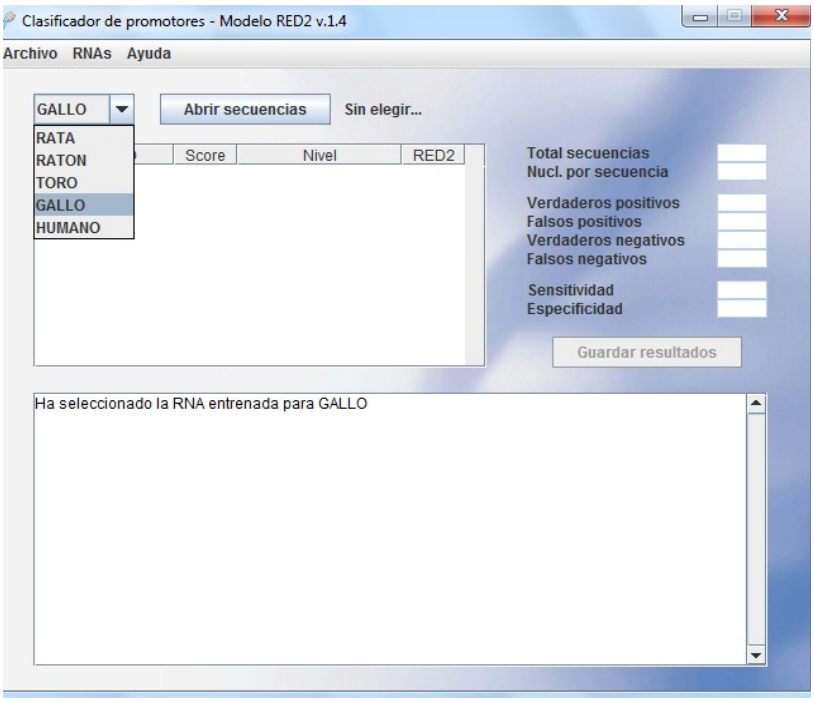


Figura 1. Selección de la RNA.

Para cargar un archivo fasta se debe hacer clic en el menú "Archivo", opción "Abrir" o desde el botón "Abrir secuencias". Aparecerá un cuadro de diálogo donde se puede seleccionar el archivo fasta que se usará como entrada a los algoritmos de clasificación y/o predicción. En la Figura 2 se muestra la carga del archivo "Gallo-entrenamiento-420-210.txt".

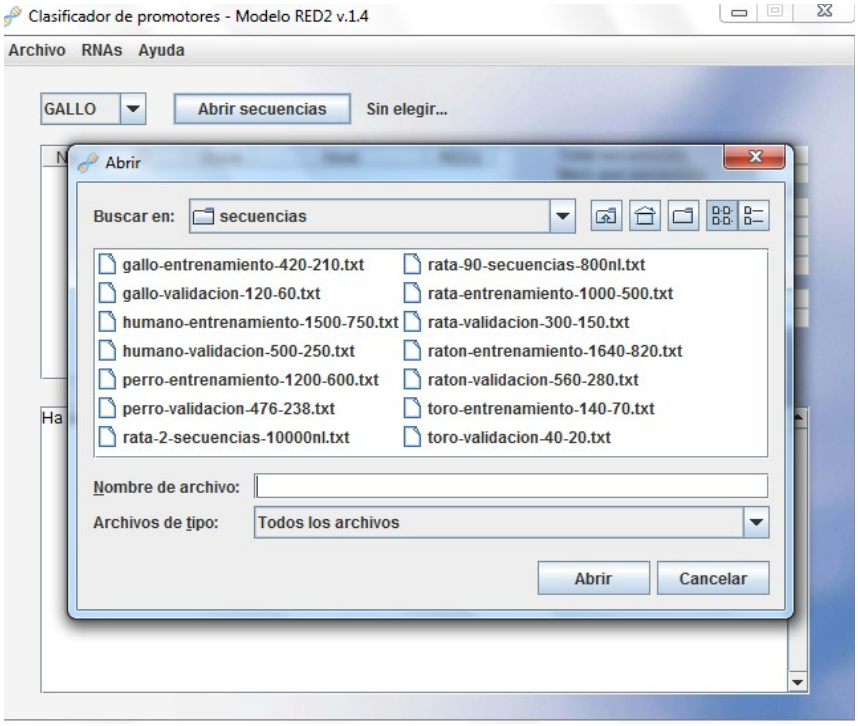


Figura 2. Selección del archivo fasta.

Inmediatamente después de seleccionar el archivo se preguntará al usuario por la cantidad de secuencias positivas, ver Figura 3. Si se desconoce se ingresará el valor de 0 "cero" seguido del botón "aceptar" o click en el botón "cancelar" si no se digita nada; en caso de conocer la cantidad de secuencias positivas la aplicación, además de clasificar o predecir promotores, calculará los indicadores de exactitud de especificidad y sensitividad,

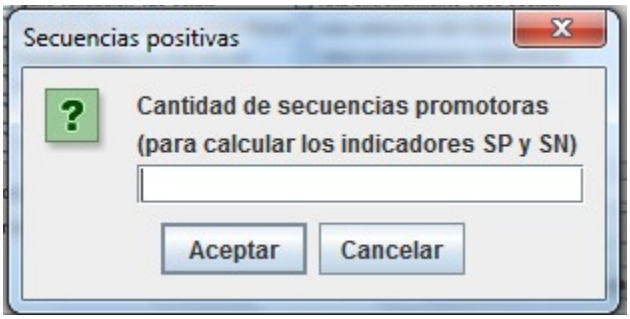


Figura 3. Ingreso de parámetro

Después de la pantalla anterior la aplicación procederá a buscar promotores en las secuencias candidatas. Para nuestro ejemplo, como el archivo de entrada contenía secuencias de 250 nucleótidos de longitud, la aplicación procederá a clasificar cada secuencia, donde 1 "uno" significa que es promotor y 0 "cero" no-promotor, ver Figura 4.

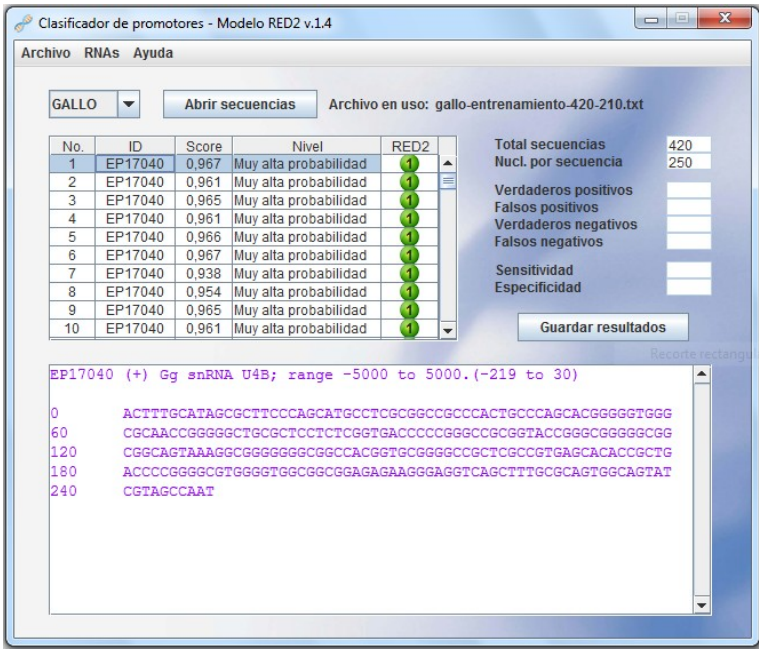

Clasificador de promotores - Modelo RED2 v.1.4
Archivo RNAs Ayuda
GALLO
Abrir secuencias
Archivo en uso: gallo-entrenamiento-420-210.txt
No. ID Score Nivel RED2
1 EP17040 0,967 Muy alta probabilidad 1
2 EP17040 0,961 Muy alta probabilidad 1
3 EP17040 0,965 Muy alta probabilidad 1
4 EP17040 0,961 Muy alta probabilidad 1
5 EP17040 0,966 Muy alta probabilidad 1
6 EP17040 0,967 Muy alta probabilidad 1
7 EP17040 0,938 Muy alta probabilidad 1
8 EP17040 0,954 Muy alta probabilidad 1
9 EP17040 0,965 Muy alta probabilidad 1
10 EP17040 0,961 Muy alta probabilidad 1
Total secuencias 420
Nucl. por secuencia 250
Verdaderos positivos
Falsos positivos
Verdaderos negativos
Falsos negativos
Sensitividad
Especificidad
Guardar resultados
EP17040 (+) Gg snRNA U4B; range -5000 to 5000.(-219 to 30)
0 ACTTTGCATAGCGCTTCCCAGCATGCCTCGCGGCCGCCCACTGCCCAGCACGGGGGTGGG
60 CGCAACCGGGGGCTGCGCTCCTCTCGGTGACCCCCGGGCCGCGGTACCGGGCGGGGGCGG
120 CGGCAGTAAAGGCGGGGGGGCGGCCACGGTGCGGGGCCGCTCGCCGTGAGCACACCGCTG
180 ACCCCGGGGCGTGGGGTGGCGGCGGAGAGAAGGGAGGTCAGCTTTGCGCAGTGGCAGTAT
240 CGTAGCCAAT

Figura 4. Resultados de la clasificación de las secuencias candidatas.

Los resultados se pueden guardar en un archivo de texto plano haciendo clic desde el menú "Archivo" seguido de "Guardar resultados" o directamente desde el botón "Guardar resultados". Aparecerá un cuadro de diálogo donde se permite elegir la ubicación y nombre del archivo donde se desea guardar los análisis, ver Figura 5.

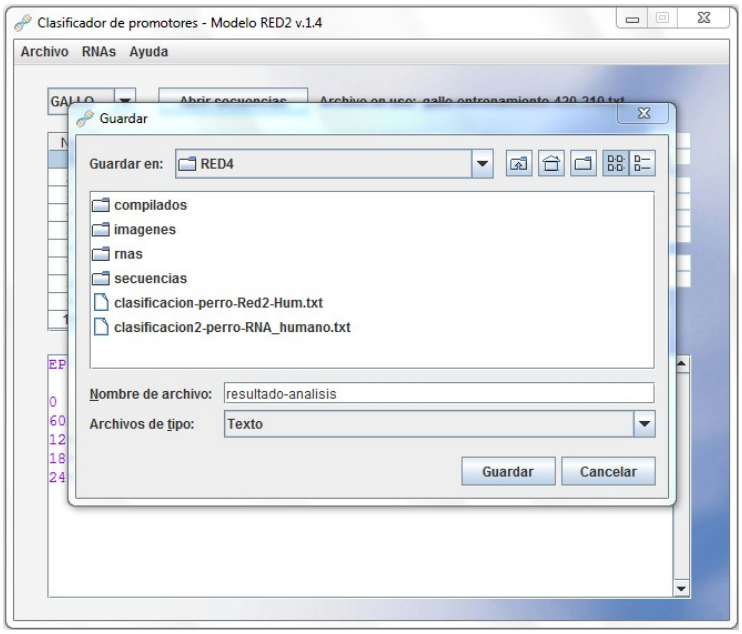
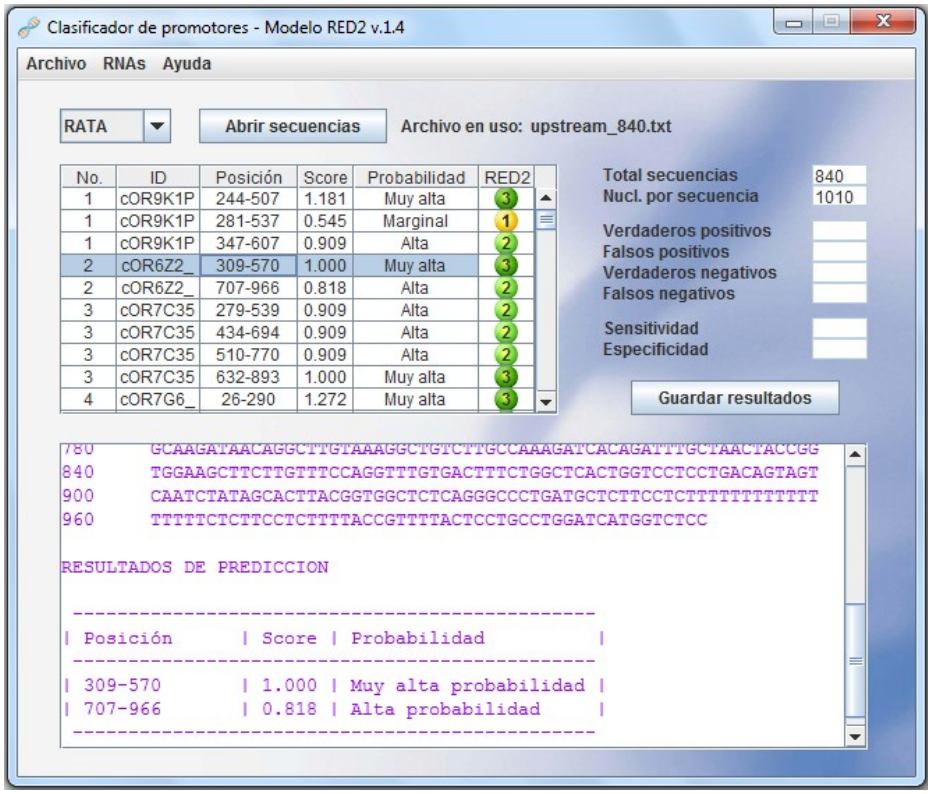


Figura 5. Guardando los resultados.

Si el archivo de entrada contiene secuencias con tamaños mayores de 250 nucleótidos, la aplicación procederá a predecir promotores, donde 3 “tres” significa muy alta probabilidad, 2 “dos” alta probabilidad, 1 “uno” marginal probabilidad y 0 “cero” no encontró, ver Figura 6.



No.	ID	Posición	Score	Probabilidad	RED2
1	cOR9K1P	244-507	1.181	Muy alta	3
1	cOR9K1P	281-537	0.545	Marginal	1
1	cOR9K1P	347-607	0.909	Alta	2
2	cOR6Z2_	309-570	1.000	Muy alta	3
2	cOR6Z2_	707-966	0.818	Alta	2
3	cOR7C35	279-539	0.909	Alta	2
3	cOR7C35	434-694	0.909	Alta	2
3	cOR7C35	510-770	0.909	Alta	2
3	cOR7C35	632-893	1.000	Muy alta	3
4	cOR7G6_	26-290	1.272	Muy alta	3

Figura 6. Resultados de la predicción de las secuencias candidatas.